Towards Understanding Complex Known-Item Requests on Reddit

Meier, Florian Maximilian; Bogers, Toine; Gäde, Maria; Thomsen, Line Ebdrup

Published in:
Proceedings of the 32nd ACM Conference on Hypertext and Social Media (HT '21)

DOI (link to publication from Publisher):
10.1145/3465336.3475096

Publication date:
2021

Document Version
Accepted author manuscript, peer reviewed version

Citation for published version (APA):
Meier, F. M., Bogers, T., Gäde, M., & Thomsen, L. E. (2021). Towards Understanding Complex Known-Item Requests on Reddit. In O. Conlan (Ed.), *Proceedings of the 32nd ACM Conference on Hypertext and Social Media (HT '21)* (pp. 143-154). Association for Computing Machinery. https://doi.org/10.1145/3465336.3475096

Towards Understanding Complex Known-Item Requests on Reddit

Florian Meier
Aalborg University Copenhagen
Copenhagen, Denmark
fmeier@hum.aau.dk

Toine Bogers
Aalborg University Copenhagen
Copenhagen, Denmark
toine@hum.aau.dk

Maria Gäde
Humboldt-Universität zu Berlin
Berlin, Germany
maria.gaede@ibi.hu-berlin.de

Line Ebdrup Thomsen
Aalborg University Copenhagen
Copenhagen, Denmark
let@hum.aau.dk

ABSTRACT

Given the important role of search engines in our everyday lives, a better understanding of the information needs that guide our information seeking behavior is essential. Known-item needs form a particular type of information need and occur when a user has a limited but concrete description of an existing object and would like to (re-)find it. Most studies of know-item needs have focused on the short query representations of these needs as they occur in search engine logs. In this article, we focus on richer, more complex known-item need representations posted to six dedicated Reddit discussion forums in the casual leisure domain. An analysis of 462 known-item requests from these subreddits revealed 33 different relevance aspects of items in a variety of different domains. Some of these aspects are highly domain-specific, while others are broadly applicable across domains. The domain also has a strong influence on the length of the known-item requests. Our findings can be used to prioritize efforts to help existing search engines better support known-item needs, both by highlighting which aspects are easier to classify automatically and by determining which information sources should be added to a search engine's index.

KEYWORDS

Known-item needs, information needs, re-finding, complex search, relevance, casual leisure, Reddit

ACM Reference Format:
Florian Meier, Toine Bogers, Maria Gäde, and Line Ebdrup Thomsen. 2021. Towards Understanding Complex Known-Item Requests on Reddit. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Given the increasingly important role that search engines play in our everyday lives, research into how people search for information, what they search for, and how we should design search engines to best support this behavior is essential. One particularly important area of study is examining the *information needs* that give rise to our information seeking behavior [8, 16, 57, 62]. Different types of needs may require different (meta)data sources to solve them and a better understanding of these dependencies could help improve the quality of search engine results. Different taxonomies of information needs have been proposed over the years [17, 33, 34, 61], with Broder's distinction between informational, navigational and transactional needs being one of the most influential. One particular type of information need is the so-called *known-item* need. While Broder's taxonomy focuses on the Web, a more domain-agnostic definition by Dahlström and Gunnarsson [24] describes it as a situation where a user has a limited but (mostly) correct description of an existing object. Typically, the user is convinced the object exists and would like to find it—although its actual existence is not required and can be doubted even by the user. This is closely related but somewhat different to re-finding, where users try to get back to previously viewed or possessed information that they know to exist [35]. Recently Arguello et al. introduced the *tip of the tongue* information need, which describes a type of known-item need for which the searcher is lacking words to formalize a concrete query.

Estimates of the frequency of know-item needs vary by study and by domain: for the Web, Pass et al. [51] estimated around 21% of queries in a log analysis study to be navigational (or know-item) searches, whereas Teevan et al. [63] showed that as many as 40% of web search engine queries are dedicated to re-finding previously visited websites. A study of digital library catalogs by Chapman et al. [22] revealed 44% of all queries to represent know-item needs and a study of classical OPAC usage by Behnert and Lewandowski [6] showed that only 56% of all known-item requests were successful. Their analysis of unsuccessful sessions revealed that in roughly 40% of all cases, queries either contained incorrect data or were too ambiguous.

Taylor [62, p. 392] described how information needs can exist at four different levels: (1) *visceral*, which represents the actual but unexpressed need for information; (2) *conscious*, where the user creates a mental description of it; (3) *formalized*, which is the "qualified and rational" statement; and (4) *compromised*, where the formalized need has been constrained and transformed to match the input format of a specific information system. Query log studies inevitably deal with the latter category of compromised need

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

representations, which means that important details could be lost in the transformation.

In this paper, we describe the results of an analysis of known-item needs posted to dedicated forums on Reddit, a popular discussion and social news website.[1] The threads in these forums contain requests that represent real-world known-item needs in a formalized state and are, for the most part, dedicated to casual leisure situations [28]. With the free-form nature of discussions on Reddit, some of these requests even resemble what Taylor called “ambiguous and rambling” statements [62, p. 392] and, as such, are close to his concept of a conscious information need. We analyze the different domains represented by these known-item needs and code them for the relevance aspects expressed in them—which properties (and context factors like previous experiences) of the known item that the user wants to find, make that particular object relevant to their need? Moreover, several users describe the earlier steps they already took to solve their known-item needs. In many cases, they explicitly state that they were unable to address their known-item needs using existing search engines and have therefore turned to Reddit. This is similar to the findings by Morris [48], who also found that people turn to social media when search engines fail them. Earlier work by [13] estimated that LibraryThing alone contained 25,000 complex needs, with many more websites such as IMDB and Reddit containing similar requests. Previous work has shown that the share of known-item needs in these requests to range from 30% to 85%, which suggest the presence of hundreds of thousands of such known-item requests on the Web. Given that many Web users are unlikely to be aware of the possibility to post such questions to Web fora, there are potentially a magnitude more of complex known-item that are going unmet by search engines. Our main contribution with this paper an analysis of the variety that make such known-item needs too complex for Web search engines to solve and is thus a first step towards a more detailed understanding of how to design search systems to better deal with these complex needs.

2 RELATED WORK

In the past, we have seen a commensurate increase in research into how people search for information, what they search for, and how we should design search engines to best support this behavior. Understanding the *information needs* that give rise to our information seeking behavior is of special importance here [8, 16, 20, 57, 58, 62]. According to Taylor, the assessment of information needs can vary radically from person to person and change with each step taken by an individual user. Taylor considers four types to represent different stages in the development of an information need, although others have proposed that they are not necessarily sequential stages, but rather different levels of understanding of the same need [23]. This allows users to make predictions about future information needs, which is essential when they evaluate keeping decisions for easier future re-finding of information [19].

The most frequently used definition of an information need is the known-item need. Known-item needs are common in many different domains and are typically (re-)defined in relation to those domains. Dahlström and Gunnarsson [24] define a known-item need as a situation where a user has a limited but (mostly) correct description of an existing object.[2] Recently, Arguello et al. coined the term, *tip of the tongue* information need following the phenomenon described by Brown and McNeill where speakers find themselves in "a state in which one cannot quite recall a familiar word but can recall words of similar form and meaning" [18]. Arguello et al. define that need as "an item identification task where the searcher has previously experienced or consumed the item but cannot recall a reliable identifier" [2, p.5]. This is very close to Taylor's description of a conscious need lacking the formalized state. However, despite its influence on the field of information science, Taylor's work has mostly been used at a conceptual level and also Arguello et al. do not consider Taylor's classification further. Ruthven performed one of the first empirical investigations of Taylor's information need classification by analyzing the linguistic differences between conscious and formalized needs present in 1,149 posts from four UK-based Internet forums devoted to different areas of life: diabetes, finance, mothers, and sexuality. The work we present in this article is similar to that of Ruthven [54] in that we also analyze information needs posted to dedicated Internet discussion forums. While some of the collected needs resemble conscious information needs—often in the form of what Taylor called “ambiguous and rambling” statements [62, p. 392]—most of them are formalized representations of the users' known-item needs. The nature of the Reddit discussion groups that we have targeted is likely responsible for this. While each domain has information need types unique to that domain, known-item needs appear to exist across domains. In our analysis of information needs posted to Reddit forums, our focus is exclusively on known-item needs in the casual leisure domain.

2.1 Re-finding needs & re-finding behaviour

A specific subset of known-item needs are re-finding needs. While in known-item search scenarios, users can make use of meta or context information about an information item they *believe* to exist, re-finding needs are characterized by the fact that users try to get back to previously viewed or possessed information that they *know* to exist [35].

A vast body of work shows that re-finding is a core information seeking behavior and has been studied in many contexts within and across different applications [10] and types of information such as physical paper documents [45], websites [49, 55, 63], emails [1, 26, 32, 66], social media posts [47], books [39, 41], games [12, 36], photos and images [9, 53], motion pictures [11, 31, 67] and music [4, 37, 43]. Researchers often discuss and list known-item and re-finding needs separately. However, we argue that these are closely related needs that evoke similar behaviors. To the best of our knowledge, there is no work that discusses how these two needs relate, even though a clear definition of what characterizes re-finding needs and behaviors does exist. Capra [21] highlighted the differences between re-finding and general search: (1) prior experience, such as different levels of knowledge in relation to the target information (known item), and different levels of expectation to satisfy the need (exact match); and (2) prior frequency, as in the

[1] Available at http://reddit.com, last visited July 21, 2021.

[2] For a more in-depth discussion and comparison of different definitions of known-item search, we refer the reader to the work by Lee et al. [44].

frequency of the task in the same session. As users have found and possibly interacted with an information item before, they do not only have knowledge about general metadata of that item (e.g., author), but also additional information from their first encounter. Studies investigating re-finding behavior show that users remember multiple aspects about the item, e.g., how, where, and when they found it, i.e., which strategy they employed, as well as time and context of the first exposure [5, 10, 25]. Previous exposure, however, is not a guarantee for successful re-finding behavior. Memories fade and information environments (e.g., email inbox, the Web, social media streams) change, which is why re-finding can be a very difficult task [27]. Capra [21] addresses the difficulty of different re-finding tasks (as summarized by Sadeghi et al. [55]). Studies that investigate aspects that are associated with difficulty in re-finding or the frustration with which users are left behind when it is unsuccessful have been performed by [27] or [46].

2.2 Richer representations of known-item needs

From a system-centered perspective, known-item retrieval has received a fair degree of attention, ranging from work on constructing test beds for the automatic evaluation of known-item retrieval in the Cranfield paradigm [3, 7] to work on optimizing and evaluating retrieval algorithms for known-item search [50, 60]. Recently, there has been more work on studying richer representations of known-item needs posted to online discussion forums and social question-answering websites. Among the first to take such an approach were Hagen et al. [31], who collected a large number of known-item questions (and answers) posted to social question-answering website Yahoo! Answers[3] from three different domains: websites, movies, and music (focusing on either songs and albums as known items). Their end goal was to create a Web-scale test collection for known-item search, so they filtered out all unsolved questions and all questions where the relevant known item was not present in the ClueWeb09 collection[4] with a comprehensive page covering that entity. While the work by Hagen et al. overlaps with our own in its focus on collecting and analyzing known-item needs posted to online discussion forums, there are also many significant differences. One of the main differences is our much broader focus by considering all the domains expressed in the known-item needs as opposed to only movies, music, and websites. To aid in the development of better retrieval and recommendation algorithms that can identify relevant books automatically [14, 15, 38], later work on these book requests focused on more in-depth analysis of the *relevance aspects* expressed by users, i.e., which aspects of the desired books make them relevant to the user—with a focus on cognitive relevance as defined by Saracevic [56]. Originally, Koolen et al. adapted the seven main relevance aspects identified by Reuter [52] in her study and extended it with known-item needs, but later Bogers et al. revisited and expanded upon this coding scheme for relevance aspects [11]. In addition to book requests on LibraryThing, other domains that have been analyzed for (among others) known-item needs include the IMDB fora for movies [11] and dedicated gaming subreddits on Reddit [12]. The earlier mentioned study by Arguello et al. analyzed questions (i.e., information requests) posted to the community question answer site *I Remember This Movie*[5]. Their qualitative coding of a set of information needs related to movies indicates that searchers employ a variety of information-seeking strategies, including semantic and episodic memories of previous experiences with the item. Additionally, they perform retrieval experiments that showed that there is substantial room for improving systems to support this kind of requests [2]. Our contribution can be seen as a generalization and extension of Arguello et al.'s work as our analysis is not limited to movie requests but covers requests from many domains that occur on these subreddits. This will increase awareness regarding for which domains complex known-item requests are most prevalent and which relevance criteria mentioned by users are shared across domains and which ones are domain-specific.

To sum up, it has been shown that complex information needs pose special challenges for search and recommendation systems and are one of the reasons people turn to social media or other community question answering platforms, like the Reddit forums dedicated to known-item requests we study in this paper [2, 30, 40, 54]. Through our analysis, we hope to contribute to a better understanding of the complexity of information needs.

3 METHODOLOGY

In order to obtain rich descriptions of information needs, or what Taylor refers to as formalized needs—"qualified and rational" statements of need [62]—we turn to Internet discussion forums. Previous work on complex search has shown that such forums are a fruitful source of complex information need representations [11, 12, 31, 39].

3.1 Data collection

Reddit contains over 2.1 million so-called *subreddits*, discussion groups covering a huge variety of topics.[6] A small number of these subreddits are dedicated to sharing and answering known-item needs. For this study, we collected data from six different subreddits. One of these subreddits, /r/tipofmytongue, is a forum containing known-item requests for all domains and is the most popular of the six subreddits with 407,044 subscribers at the crawling time. The other five are domain-specific: games (/r/tipofmyjoystick), books (/r/whatsthatbook and /r/namethatbook), and music (/r/NameThatSong and /r/whatsthatsong). Initially, we considered focusing only on the /r/tipofmytongue subreddit for a cleaner focus, but with several other subreddits dedicated to specific domains, we were afraid it would harm the representativeness of the different domains in our sample. By including them, we hope to counteract the fragmented nature of some of these domains and get a more representative snapshot of complex domain-specific known-item needs shared on Reddit. For instance, there are more game- and book-related known-item needs in the dedicated subreddits than in the catch-all /r/tipofmytongue subreddit.

[3] Available at https://answers.yahoo.com/, last accessed July 21, 2021
[4] Available at https://lemurproject.org/clueweb09.php/, last accessed July 21, 2021
[5] https://irememberthismovie.com/, last visited July 21, 2021
[6] According to http://redditmetrics.com/, last visited July 23, 2020.

We adapted an existing Reddit crawler[7] to collect known-item requests from the six subreddits. Each thread consists of a first post where a user expresses their known-item need and any number of comments from other Reddit users attempting to answer the user's question. We continuously crawled all threads and comments posted to these six subreddits from June 2-22, 2018, resulting in 9,845 threads and 37,189 comments. About 20% (n = 1,932) of the 9,845 threads did not receive a single reply. Table 1 shows an overview of these threads in terms of thread and comment activity, number of subscribers, average request length, and use of URL linking. It also shows the number of threads taken from each subreddit in our coding sample of 462 threads (third column).

Reddit allows subreddit moderators to include posting guidelines and pin them to the top or side of the subreddit so that they are always visible. In addition to containing rules about what content gets users banned from the subreddit, these guidelines can also contain recommendations on how best to phrase a request to increase the chances of it being fulfilled. For instance, /r/tipofmyjoystick recommends users include information about the platform(s), genre, estimated year of release, graphics/art style, notable character, notable gameplay mechanics, and any other details the user can remember. Subreddits marked with a ⋆ in Table 1 contain pinned guidelines on how to write effective posts that make re-finding easier. We examine the potential effect of these guidelines on the requests in Section 4.

3.2 Coding

Inspired by Schreier's toolbox for qualitative content analysis [59], we chose a three-stage coding process consisting of *Open coding*, *Axial coding*, and *Final (Selective) coding* as our basic strategy for inductively creating a coding scheme.

3.2.1 Open coding. To analyze the known-item requests posted in the six subreddits, we developed a coding scheme for the information expressed by the requesters using open coding. We selected a sample of 100 threads, randomly sampled from the entire data set, to serve as development set. Three of the authors then individually developed their own set of codes on the same sample. For each thread, coders were shown the title and the full text of the first post containing the original known-item requests; the subreddit it originated from was not included to avoid any biases.

3.2.2 Axial coding. Open coding resulted in three different sets of codes with a combined total of 264. Examining these codes revealed two broad types of codes: (1) codes related to the *domain* (e.g., books, movies, TV shows, video games), and (2) codes related to the *relevance aspects* of the known-item, i.e. the properties that the user remembers about the item that make that known item relevant and other items not relevant. In the axial coding phase, we therefore produced two different coding schemes: one for the domain and one for the relevance aspects. The majority of the 264 codes were proposed by multiple annotators and we used affinity diagramming to identify relationships between codes and re-arrange them into higher-level categories. When grouping and merging codes together, we tried to keep in mind the ultimate goal of designing information search systems that can help answer known-item needs. This meant, for instance, that if we expected different relevance aspect codes to be covered by the same type of metadata or information source, we combined them into a single concept. For example, requests that reference the dialogue in a movie or the lyrics in a song were originally coded as separate concepts, but later combined as the sub-category Dialogue & lyrics[8]. Some codes were further divided into sub-categories or concepts when the distinction was deemed useful, e.g., when Musical properties was subdivided into Instruments, Melody, Rhythm & tempo, and Sound & effects. All four authors discussed the resulting coding schemes until consensus was reached about the categories and their labels. Textual descriptions and aspects along with prototypical examples were added to aid the final annotation process. Our final coding scheme is shown in Figure 1; see Section 4.1 for further details[9].

3.2.3 Final coding. After calibrating our two coding schemes, each coder was assigned 140 threads, randomly selected from the six subreddits for the final coding phase. Posts used in the development set were not re-used in this final sample. Not every subreddit thread was necessarily a known-item request; such posts were deleted from the sample and not annotated. We included a total number of 178 posts that overlapped between multiple annotators in order to examine reliability. We used Fleiss' kappa to calculate coding agreement for these posts, the results of which are described in more detail in Section 4.2. After removing these multiple-assessed duplicates and non-known-item requests, we ended up with a total of 462 unique annotated threads. After the first coding round, all annotators discussed their experiences, which led to small refinements of the category labels and descriptions, as well as the addition of the sub-category Physical properties. Each annotator then revisited their requests to adjust their annotations.

4 REQUESTS

In this section, we present our analysis of the known-item requests. We start by introducing our coding schemes, followed by information about inter-annotator agreement in Section 4.2. Finally, Section 4.3 presents a detailed analysis of the complex known-item requests in our sample.

4.1 Coding schemes

4.1.1 Domains. Our first coding scheme focused on the domain represented by the known item. All 462 known-item requests were assigned to one or more domains; most requests were assigned to a single domain with 20 requests falling under two domains. The most frequently covered domains were Music at 32.3% of our sample (e.g., *"I remember the song starts off with multiple female vocals singing together in unison but they aren't singing any lyrics"*), Game at 17.5% (e.g., *"Horror game that is in asylum and there is a guy that is like a cripple with an axe"*), and Motion picture at 13.2% (e.g., *"Movie about a young boy who's kidnapped by his father from his mother after his birthday party, then eventually returned to*

[7] Available at https://github.com/lucas-tulio/simple-reddit-crawler, last visited July 23, 2020.

[8] In the rest of the paper, all codes will be marked up in red (e.g., Plot) and all domains will be marked up in green (e.g., Game).

[9] Our code book along with the annotated dataset is available from http://toinebogers.com/?page_id=256.

Table 1: Statistical summary of the 9,845 known-item request threads crawled from Reddit. Subscriber counts are reported for the crawling period. Subreddits marked with ⋆ include guidelines on how to write effective posts that make re-finding easier.

Subreddit	Threads total	Threads sample	Subscribers	Comments	Avg. comments/ thread	Avg. request length (tokens)	% of requests with URL
/r/tipofmytongue⋆	7,021	329	407,044	29,734	4.2	84.6	18.6
/r/tipofmyjoystick⋆	1,131	52	38,886	4,969	4.4	124.6	6.5
/r/NameThatSong⋆	923	44	11,944	906	1.0	58.5	46.2
/r/whatsthatbook⋆	713	31	12,977	1,521	2.1	134.2	1.3
/r/namethatbook⋆	29	3	999	46	1.6	113.9	0.0
/r/whatsthatsong	28	3	131	13	0.5	68.3	21.4
Total	9,845	462	N/A	37,189	3.8	97.3	15.7

his mother years later and is all wild and unruly."). Other common domains included Book (n = 54, 11.7%), Video (n = 37, 8.0%), TV show (n = 32, 6.9%), Social media post (n = 16, 3.5%), Person (n = 13, 2.8%), and Image (n = 10, 2.2%). The remaining 28 (6.5%) posts (Other) covered many additional domains, such as users interested in re-finding toys, Web comics, laptops, and linguistic expressions.

4.1.2 Relevance aspects. Our final coding scheme for relevance aspects (see Figure 1) includes five top-level categories: Content, Metadata, Experience, Context of exposure and Search process. These five categories are further divided into 24 sub-categories from which some are specified with concepts. Again, none of these aspects are mutually exclusive; requests could be assigned to more than one relevance category. The top bar chart in Figure 2 shows the distribution of our 33 codes over our sample of 462 threads for all domains combined, ordered by relative frequency. The rest of Figure 2 will be discussed in Section 4.3.2.

The top-level category Content covers aspects related to the question what the known item is about or contains, and 86.4% of our sample have some respective code. It consists of eight sub-categories with three of them being further subdivided into concepts. The most frequent sub-categories for Content are General plot (33.5%), Character(s) (31.2%) and Event(s) (24.5%). While most of the sub-categories are generally applicable across domains, such as Character, Dialogue & lyrics (16.7%), or General plot, other sub-categories are more domain-specific, such as Musical properties (13.4%) and Gameplay mechanics (14.7%).

The top-level category Metadata describes metadata attributes of the known-item and makes up 70.8% of our sample. It consists of nine sub-categories, with Person(s) involved further subdivided into two concepts. The majority of metadata elements used in these categories can be applied to most domains, however there are some challenges. For example, the information marked with the sub-category Release date (26.0%) shows a wide range of specificity, from specific release dates to deducing a multi-year release window based on vague memories of the user. Some sub-categories are specific to certain domains, such as Platform (19.0%), which usually describes the platform or device for Game requests. Interestingly, users pointed often to information that is usually not covered through metadata elements such as Gender (12.1%) with respect to voices within Music requests or of characters within Motion picture.

About 12.1% of our sample comprises known-item requests that include information about the kind of Experience the item provided. These include the Mood (8.4%) the item evoked in the user (e.g., *"Inappropriate/cringe game show moment where contestants had to guess an image based off of squares that were slowly revealed"*) as well as the Popularity (3.9%) or obscurity of the item (e.g., *"I think that the video had plenty of views to find easily, but i can't find it myself"*).

The top-level category Context of exposure—which makes up 38.7% of our sample—can be described as a Situation of exposure (25.3%) code (e.g., *"this came up in a Japanese restaurant and I asked the waitress but she didn't know what it was, ..."*) or as a Time of exposure (21.6%) code (e.g., *"I was maybe 6-9yrs old, born in '92"*).

Finally, Search process—36.8% of our sample—is comprised of five sub-categories that describe the user's efforts to include information they believe may aid in the re-finding process. One example is History (9.1%), where they describe their previous efforts to re-find the known-item (e.g., *"I spent an hour Googling it about a month ago because it was stuck in my head, insanely catchy melody. I finally came across it, but Shazam and other apps have no idea what it is."*). Some users provide examples Similar to the known item (11.7%) or explicitly rule out other items (Not this one, 5.6%). Others include additional, non-textual information by linking to an External resource (15.8%, e.g., *"[LINK TO YOUTUBE] What is the song at 3:45?"*) or to an Own recording (4.3%, e.g., *"What was this song? This is what the melody was like! <LINK TO RECORDING>"*).

4.2 Agreement

To assess the inter-annotator agreement between coders, we arranged for 178 posts to overlap with between pairs of coders. Our coding scheme is not mutually exclusively and due to our use of multiple coders with only pairs of coders sharing overlap, we used Fleiss' kappa calculated this for each code separately [29]. Following the guidelines by Landis and Koch [42], we then took the average of kappa values for each single code for an overall $\kappa_{overall}$ of 0.57, which corresponds to moderate agreement. Simple non-chance-corrected agreement was 93% on average over all codes.

There is a clear difference between the five top-level categories in terms of agreement and, by proxy, how difficult they are to code for. Search process had the highest agreement with κ = 0.73, with Content second with κ = 0.50. Coding for Metadata resulted in κ = 0.45; Context of exposure yielded κ = 0.43 and Experience

Top-level category		Sub-category	Concept	Description	n	%
What was the item about?	Content	Character(s)		Item with specific characters, types of characters, or character development	144	31.2
		Design		Item with a particular art style, physical design, or graphics	71	15.4
		Dialogue & lyrics		Item containing specific dialogue, quotes, or lyrics	77	16.7
		Plot	General plot	General plot or narrative of the item	155	33.5
			Event(s)	One or more specific events happening and/or when they occur in the plot	113	24.5
		Setting	Location	Item taking place in a specific physical setting, location or near landmarks	36	7.8
			Time period	Item set in a particular time period or around a specific historical event	10	2.2
		Musical properties	Instruments	Describing an item by specific instruments used in it	29	6.3
			Melody	Description of the melody of the item	25	5.4
			Rhythm & tempo	Description of the item's rhythm and/or tempo	18	3.9
			Sound & effects	Describing the generic sound or effects used in the item	24	5.2
		Gameplay mechanics		Gameplay mechanics or functionality of the item	68	14.7
		Topic		Item covering or addressing specific topic(s)	43	9.3
What kind of properties it have?	Metadata	Cultural properties		Cultural properties of an item (e.g., language)	42	9.1
		Gender		Gender of the people involved in the item	56	12.1
		Genre		Genre of an item	169	36.6
		Length		Details about the items length, size, or duration	12	2.6
		Person(s) involved	Creator(s)	Creator(s) of the item (e.g., directory, writer, composer)	23	5.0
			Contributor(s)	Contributors to the item (e.g., singers, musicians, (voice) actors)	25	5.4
		Physical properties		Physical properties of the item (e.g., cover, packaging, material)	26	5.6
		Platform		Platform the item runs on or was broadcast on	88	19.0
		Release date		The original release date/period of the item (or estimate of this)	120	26.0
		Series		Series or franchise the item a part of	29	6.3
		Title		Describing (part of) the title of the item	19	4.1
What experience did it provide?	Experience	Mood		The mood, tone or feeling evoked by the item	39	8.4
		Popularity		The popularity or obscurity of the item	18	3.9
How was it originally encountered?	Context of exposure	Situation of exposure		Description of the situation/setting in which the user first encountered the item	117	25.3
		Time of exposure		Description of the time period in which the user first encountered the item	100	21.6
What could help locate the item?	Search process	Examples	Similar to	Support re-finding by providing examples of items that are similar in some way	54	11.7
			Not this one	Support re-finding by explicitly ruling out candidate items as the right answer	26	5.6
		Link to resource	Own recording	Support re-finding by providing a recording imitating aspects of the item	20	4.3
			External resource	Support re-finding by providing a link to an external resource with information	73	15.8
		History		Support re-finding by descring the previous steps taken by the user	42	9.1

Figure 1: The coding scheme for relevance aspects expressed in known-item requests (N = 462) along with their absolute (n) and relative frequencies (%).

resulted in the lowest agreement scores with $\kappa = 0.03$. This suggests that more subjective and contextual relevance aspects could be harder to detect and code for.

4.3 Analysis

4.3.1 Sample representativeness. Our coded set of 462 known-item requests was randomly sampled from a larger set of 9,845 threads—itself also a sample from Reddit. Table 1 summarizes this data with the third column showing the number of threads sampled from each subreddit. Our sample appears to be representative of the larger data set. The average number of comments per thread in our coded sample is close to the average in the large sample (4.3 vs. 3.8), the average request length is similar (84.7 vs. 97.3) and the percentage of requests linking to a URL are also fairly close (19.0% vs. 15.7%). The fact that our coded sample has slightly higher numbers can be

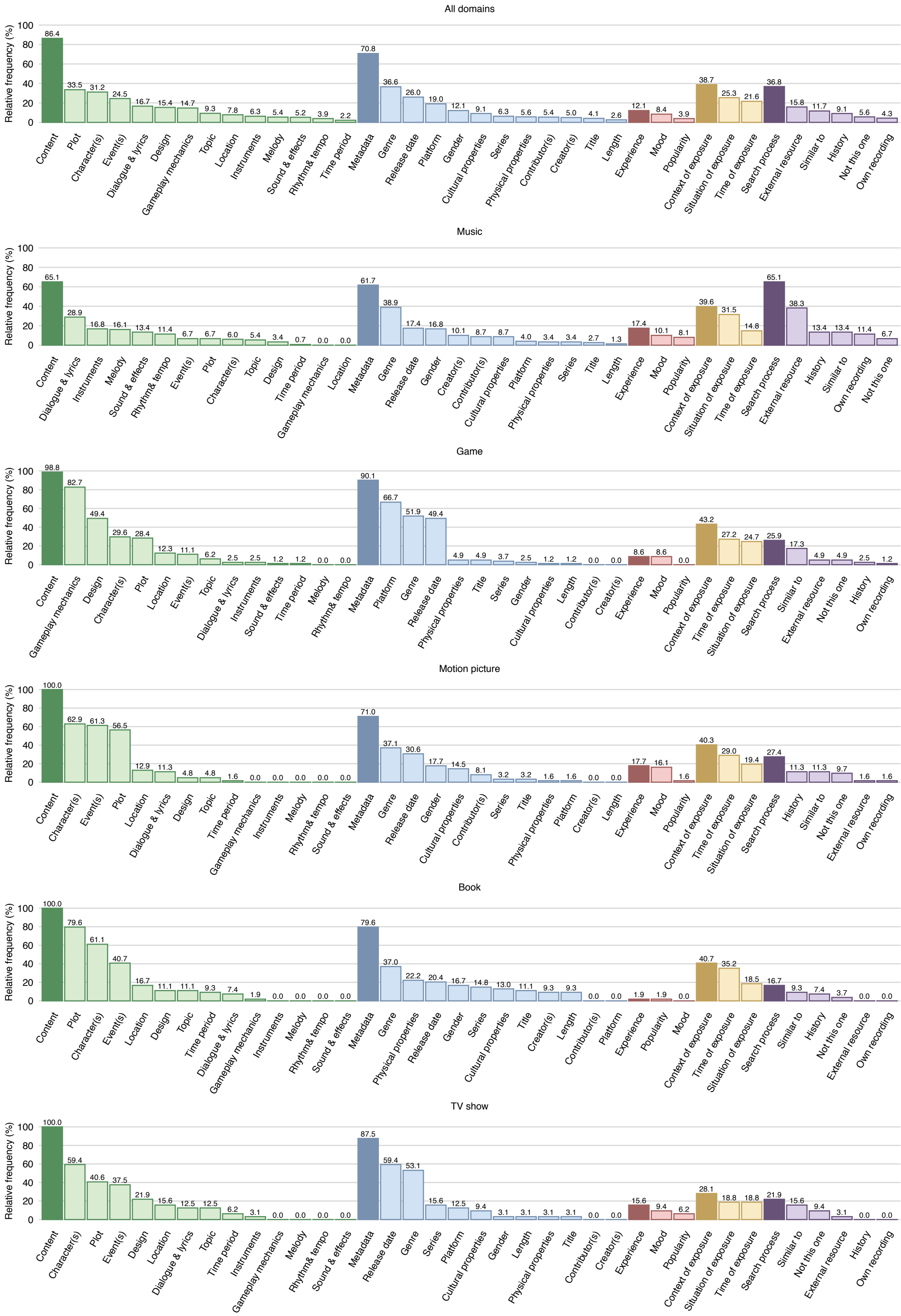

Figure 2: Relative frequencies of the 33 different codes in our sample of annotated Reddit threads ($N = 462$) for all domains as well as the top-five domains (Music, Game, Motion picture, Book, and TV show). The dark bars in each color group represents the top-level code (e.g., Content) and any lighter-colored bars to their right represent the low-level codes (e.g., Design, Gameplay mechanics).

explained by the removal of the small percentage of non-known-item requests from the sample; these were not removed from the larger data set, which resulted in the discrepancies.

4.3.2 Domain-specificity of relevance aspects. An interesting question to ask is whether any relevance aspects or top-level categories are domain-specific. Figure 3 shows how the five top-level categories are distributed over the five most popular domains while Figure 2 shows how the low-level relevance aspects are divided across the five top-level domains. Both figures reveal noticeable differences in the way relevance aspects are used to re-find items from different domains. Some domains show relatively similar distributions: Video and Motion picture are understandably similar, and the same goes for other narrative-driven domains, such as Book and Game. In contrast, Music stands out the most compared to the other four top-five domains: known-item requests for Music focus less on Content, perhaps because content-based description of music is much harder for novices than it is in other domains. As a result, we see a higher proportion of Search process aspects due to the common strategy of recording part of the melody or linking to an external resource, such as a YouTube video. In contrast, Content is present in virtually all of the known-item requests from the other four domains. Metadata occurs in 70.8% of all known-item requests, and is least common for Music—due to fewer applicable metadata properties—and most common for Game requests. This is probably due to the fact that the recommended posting template for the /r/tipofmyjoystick subreddit includes metadata fields such as Genre, Platform, and Release data.

Finally, known-item needs in the Book domain are rarely described using Experience aspects—perhaps because reading a book is a more personal experience than for the other domains and therefore less likely to aid in re-finding. In general though, Experience aspects are the least frequently expressed aspects. We hypothesize that this is because users recognize that their subjective personal experiences are less helpful than describing objective properties. Subjective codes—Mood and Popularity—make up two of our 33 codes (or 6.1%), so to test this hypothesis we check whether the share of subjective codes among all codes for each request is significantly different in our sample. With a sample mean share of subjective codes at 2.66%, we calculated a confidence interval for this unknown population parameter using bootstrapping on our sample data for 10,000 iterations. The results indicate that we can be 95% confident that the true share of subjective codes is between 2.00% and 3.42%, which means there is evidence that subjective codes are used less often than expected. We also detected strong associations between domains and sub-categories. To this end, we calculated point-wise mutual information (PMI), a symmetric measure of association strength, between sub-categories and domains. Some of the most strongly associated pairs include Game-Gameplay mechanics (PMI = 7.86), Game-Platform (PMI = 7.39), Book-Physical properties (PMI = 7.51), Book-Length (PMI = 7.41), Image-Design (PMI = 7.09), Person-Gender (PMI = 7.07), and TV show-Series (PMI = 7.05), These patterns are also evident in Figure 2. Finally, the various musical property aspects are also strongly associated with the Music domain as well as the more domain-specific Own recording and External resource aspects.

4.3.3 Aspect co-occurrence. In general, the known-item requests posted to the six subreddits are complex in nature: on average, 4.1 different codes were applied to requests with; two requests even containing 10 different relevance aspects. Because of the richness of the requests, it is plausible to expect co-occurrence of different aspects: some of them are more likely to be applied to the same request than others. We therefore analyzed how often the aspects from the top-level categories co-occurred, the results of which are visualized in Figure 4. It shows that Content and Metadata are by far the most frequently co-occurring top-level categories. One reason for this is that they are also the two most frequently occurring individual categories—users most often tend to remember what an item was about and some of its properties—so they are likely to co-occur frequently as well. In addition, users might also assume that the other top-level categories are less likely to be useful to others in identifying the object they are looking for. This means that to be able to successfully address known-item needs, a search system needs to emphasize the indexing of comprehensive and quality information about these domain aspects. This requires that additional information about, for example the general plot of a motion picture and descriptions of characters appearing in it are readily available for indexing.

We also found interesting patterns in terms of how the individual sub-categories are associated with each other in terms of PMI. We find that Melody and Own recording are the most strongly associated aspects with a PMI of 8.14, which is to be expected since information about the melody of a song can be hard to describe in words. Humming or playing a few bars of a song and making this recording available is a more productive strategy in this case. In general, the four Musical properties codes are most strongly associated with each other with PMI values ranging from 7.77 to 8.08, which is unsurprising given their strong link to the Music domain. Likewise, Gameplay mechanics and Platform are strongly associated (PMI = 7.45) because they are nearly exclusively used to describe video games (Game). Other strong pairs are Location and Time (PMI = 7.77), and Situation of exposure and Time of exposure (PMI = 6.54), which confirms grouping these pairs together was a good decision. As mentioned earlier, known-item requests include, on average, 4.1 different relevance aspects with a maximum of 10. The domain with the highest number of expressed relevance aspects is the Game domain with an average of 5.0 aspects, followed by Book with 4.67 and TV show with 4.53. Requests from the domain Social Media post included the lowest number of expressed aspects with an average of 2.63. To test whether the item's domain has a significant influence on the number of expressed relevance markers, we performed a linear regression with domain as the independent variable and the number of included relevance aspects as the dependent variable. The Game domain was chosen as baseline for comparison (intercept) as it included the most relevance aspects on average. Following the model parameters, there is a significant effect of domain on the number of relevance markers ($F = 6.81$, $p < 0.001$), which allows us to conclude that domain very likely has an influence on the number of applied relevance aspects. A post-hoc pairwise comparison showed that its especially requests for Game and Book items that included significantly more relevance markers compared to requests of the domains Music, Other and Social Media post. Moreover, requests of domain Movie included

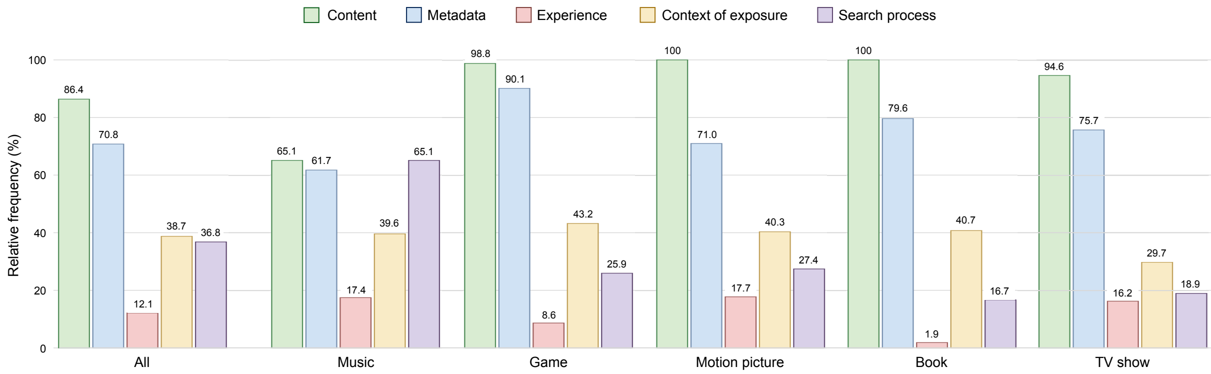

Figure 3: Distribution of top-level relevance categories over the five most popular domains for known-item requests. Vertical bars represent the relative frequency (%) of a category in all requests for that domain.

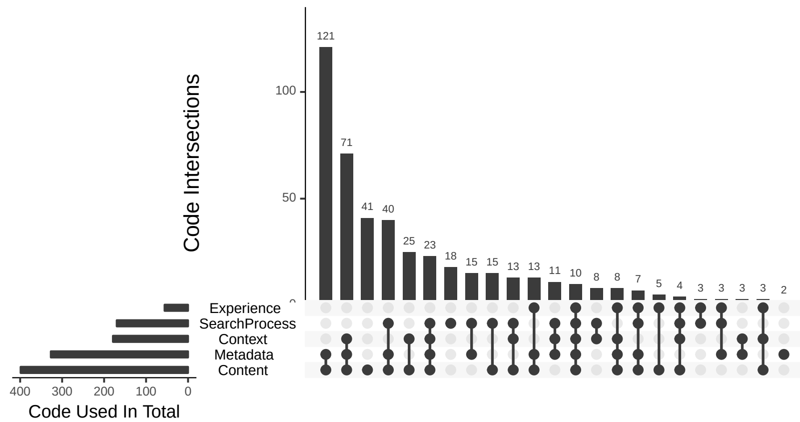

Figure 4: Visualization of the co-occurrence between top-level relevance aspect categories in known-item requests. Vertical bars represent how often the combinations of top-level aspects co-occurred in our sample (N = 462).

significantly more relevance markers than Other and Social Media post. Finally, for TV show a significant difference could be found compared to requests of Social Media post.

A related question that arises here, is whether the number of relevance aspects is related to the length of the request (in tokens). These two request characteristics do indeed correlate quite strongly ($r = 0.54$, 95% CI [0.47, 0.60], $p < 0.001$) as longer requests do include more relevance aspects. On the one hand this makes requests more complex while at the same time providing other users with more hints to solve those requests.

4.3.4 Request length. Figure 5 shows the average request length (measured in tokens) split by (a) domain, (b) top-level relevance aspect, and (c) subreddit. Figure 5a shows the average request length per domain. Four domains have above-average requests length compared to the entire sample: Book (109.3 tokens on average), Movie (107.0 tokens), TV show (102.0 tokens), and Game (96.6 tokens). What all these domains have in common, is that their duration—as in the time spent consuming items from these domains—is typically much longer than the other domains. This makes it more likely that users are able to remember more of these objects and therefore post longer requests. In contrast, describing people, songs or images is less likely to require extensive descriptions. To test this idea, we performed a linear regression with token count as the numerical outcome variable and domain as the categorical explanatory variable. The Book domain, which has the longest requests on average at 109.3 tokens, was chosen as baseline for comparison (intercept). The regression model showed a significant effect of domain on request length ($F = 6.13, p < 0.001$). Again we performed a post-hoc pairwise comparison. The estimates show that requests from the Book, Movie, Game and TV show domains are significantly longer than requests of domain Music—by 44.73, 42.40, 32.05 and 37.45 tokens on average respectively. However, these are the only significant differences.

This difference between domains is also borne out by differences between the subreddits as shown in Figure 5c: the music-focused /r/NameThatSong subreddit has an average request length of only 52.3 tokens, while /r/whatsthatbook and /r/tipofmyjoystick have average lengths of 113.8 and 103.8 tokens respectively. The /r/tipofmytongue subreddit has an average length of 82.9 tokens, which is close to the overall average of 84.7 tokens and reflects the

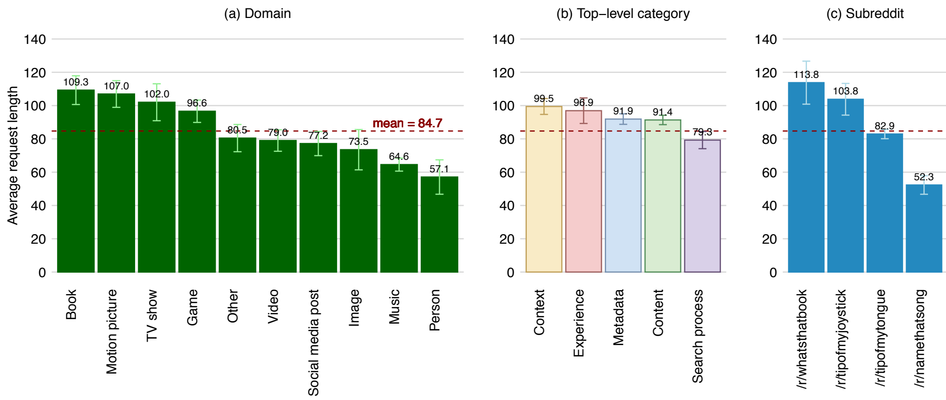

Figure 5: Visualization of average request length in tokens by (a) domain, (b) top-level category, and (c) subreddit. The average request length over all requests ($N = 462$) is denoted by the red dotted line. In Figure 5c, all subreddits with less than 10 posts have been removed.

broad nature of the forum. Figure 5b shows the average request lengths per top-level relevance category. Here, all categories except for Search process are associated with longer than average requests. Context of exposure aspects typically involve long descriptions of the original context users first encountered the object, which explains its high average request length. The below-average length of Search process requests can be explained by the fact that many of the requests tagged with this set of aspects only include links to other resources or names of other objects the user has ruled out already, thereby.

The concepts Own recording and External resource are used to mark known-item requests where the user includes a link to either their own recording or to another, external resource. Both types of links add extra information about the known item that cannot easily be expressed textually. Of all domains, the Person and Music domains have the highest percentage of posts with links at 46.2% and 45.6% respectively. As we argued before, music can be hard to describe for novices and the Person requests typically contain a link to a photo or video of the person in question. The Video and Image domains have shares of 10.8% and 10.0% respectively and in other domains less than 5% of the known-item requests come with links.

5 DISCUSSION & CONCLUSION

In this article, we have examined how people turn to online forums when their complex known-item needs cannot be solved using conventional search engines. An analysis of 462 known-item requests from casual leisure domains showed that these difficult known-item needs are complex with a wide variety of different relevance aspects expressed. These known-item requests—and online forums dedicated to collaborative search in general—make a valuable study subject, because they represent a richer, more formalized representation of the underlying information need according to Taylor's information need hierarchy. Specifically, we show that there is a great variety in the relevance aspects that users express in their known-item requests and that there is a significant difference in the complexity and length of known-item requests between different domains. We believe that our analysis of these hard known-item requests has several practical implications for the design of both domain-specific and general search engines which also align with the findings of previous studies in this context [2, 54]. One area where our findings could be of use is in deciding which metadata and other information sources should be indexed to enable better automatic retrieval for known-item needs. Aspects such as Topic or Creator are more likely to be covered by existing metadata sources, which matches their relatively low occurrence frequencies in our sample, which suggests that known-item needs that focus only on such aspects can already be solved by existing search engines. Other aspects that are more common in our sample—such as Plot, Character(s), or Gameplay mechanics—are apparently harder to deal with for existing search engines, as there are also relatively few comprehensive information sources that contain this information. As Arguello et al. note, simply indexing the required information is not enough. To properly support known-item needs, a search engine should also be able to detect the different relevance aspects in a user's query. Our analysis of inter-annotator agreement suggests that not all of these relevance categories are equally easy to detect for humans and, as a likely result, by automatic means. Knowledge of which aspects are easier to code for could also help inform search engine design by prioritizing which aspects to detect and include first. We also provide a clear overview of which aspects are important for known-item needs in different domains, making it possible for developers of domain-specific search engines to optimize their indexes for these needs.

Our findings could also help inspire different search user interfaces for dealing specifically with known-item needs. Some aspects, like Release date or Platform, could be captured by the search engine by providing additional features such as drop-down menus or sliders, whereas others could be used only for field-specific searching if the search engine interface provided a more structured user experience, analogous to the posting guidelines provided by several subreddits.

Finally, we believe that the dataset and findings we provide can be used to advance the development of conversational systems especially from an interactive information retrieval perspective

[64]. As search and assistance systems are becoming increasingly conversational in nature we see a rising demand for investigations into how to make interaction with such systems more dialog- and human-like resembling real human-to-human conversations. We believe that the requests we are analyzing can be used to improve (1) query understanding and (2) search process management. The requests are formalized representations of known-item needs and resemble in their level of detail what Trippas et al. described as *teleporting* queries, where users, when interacting with an intelligent personal assistant, describe their need in a great level of detail to move directly to the solution [64]. This way, known-item requests posted on Reddit could be used to train conversational systems to understand formalized needs [62]. Moreover, the relevance aspects we identified can be used to build dialog schemes and methods for search need elicitation. Once the domain of the item is known, the sub-level codes and categories we identified to be most strongly associated with this domain (see Section 4.3.2) could be used to further elicit information in a step-by-step conversation approach as the basis for turns in a human-machine dialog [65].

In general, we found Reddit to be a fruitful source for known-item needs representing a variety of different casual leisure domains. Other subreddits exist that focus on other types of information needs, such as subject search or recommendation by analogy. Using online data has several advantages, such the possibility of analyzing textual representations of formalized and sometimes conscious refinding needs for meaningful patterns as opposed to having to rely on query logs. Moreover, this also means we can investigate a much larger number of information needs than we could otherwise collect through interviews, diaries, and other narrative methods. However, our approach suffers from the inability to ask follow-up clarification questions which, for example, could clarify if relevance aspects mentioned in requests are due to experiences that are more likely to be shared broadly or due to experiences that are more likely to be uniquely personal. Additionally, relevance aspects added by the asker could not only be *not helpful* but factually wrong due to false memories [31]. Analysing the discussion threads following each post could potentially give insights into which aspects of a post were helpful, which needed further clarification or which could be identified by the community as being false memories.

In future work, we would like to extend our existing annotated sample and perform additional analyses of how the domains and relevance aspects interact with each other. Moreover, one should look into which relevance aspects make requests easy to be solved by the community. Automatically detecting these dimensions would be another interesting venue for future research.

REFERENCES

[1] Qingyao Ai, Susan T. Dumais, Nick Craswell, and Dan Liebling. 2017. Characterizing Email Search Using Large-Scale Behavioral Logs and Surveys. In *WWW '17: Proceedings of the 26th International Conference on World Wide Web*. ACM, Republic and Canton of Geneva, CHE, 1511–1520.

[2] Jaime Arguello, Adam Ferguson, Emery Fine, Bhaskar Mitra, Hamed Zamani, and Fernando Diaz. 2021. Tip of the Tongue Known-Item Retrieval: A Case Study in Movie Identification. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval* (Canberra ACT, Australia) *(CHIIR '21)*. Association for Computing Machinery, New York, NY, USA, 5–14. https://doi.org/10.1145/3406522.3446021

[3] Leif Azzopardi and Maarten de Rijke. 2006. Automatic Construction of Known-item Finding Test Beds. In *SIGIR '06: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, USA, 603–604.

[4] David Bainbridge, Sally Jo Cunningham, and J Stephen Downie. 2003. How people describe their music information needs: A grounded theory analysis of music queries. In *Proceedings of ISMIR '03*.

[5] D. Barreau and B. A. Nardi. 1995. Finding and Reminding: File Organization from the Desktop. *SIGCHI Bulletin* 27, 3 (1995), 39–43.

[6] Christiane Behnert and Dirk Lewandowski. 2017. Known-Item Searches Resulting in Zero Hits: Considerations for Discovery Systems. *Journal of Academic Librarianship* 43, 2 (2017), 128–134.

[7] Steven M. Beitzel, Eric C. Jensen, Abdur Chowdhury, David Grossman, Ophir Frieder, and Ophir Frieder. 2003. Using Manually-built Web Directories for Automatic Evaluation of Known-item Retrieval. In *SIGIR '03: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, USA, 373–374.

[8] N.J. Belkin, R.N. Oddy, and H.M. Brooks. 1982. ASK for Information Retrieval: Part I. Background and Theory. *Journal of Documentation* 38, 2 (1982), 61–71.

[9] O. Bergman and S. Whittaker. 2016. *The Science of Managing Our Digital Stuff*. The MIT Press, Cambridge, London.

[10] R. Boardman and M. A. Sasse. 2004. "'Stuff Goes into the Computer and Doesn't Come out"': A Cross-tool Study of Personal Information Management. In *Proceedings of CHI '04* (Vienna, Austria). ACM, New York, NY, USA, 583–590.

[11] Toine Bogers, Maria Gäde, Marijn Koolen, Vivien Petras, and Mette Skov. 2018. "What was this Movie About this Chick?" A Comparative Study of Relevance Aspects in Book and Movie Discovery. In *Proceedings of iConference 2018*. Vol. 10766. Springer, 323–334.

[12] Toine Bogers, Maria Gäde, Marijn Koolen, Vivien Petras, and Mette Skov. 2019. "Looking for an Amazing Game I Can Relax and Sink Hours into..." A Study of Relevance Aspects in Video Game Discovery. In *Proceedings of iConference 2019*. Vol. 11420. Springer, 503–515.

[13] Toine Bogers and Marijn Koolen. 2017. Defining and Supporting Narrative-driven Recommendation. In *RecSys '17: Proceedings of the Eleventh ACM Conference on Recommender Systems*. ACM, 238–242.

[14] Toine Bogers and Vivien Petras. 2015. Tagging vs. Controlled Vocabulary: Which is More Helpful for Book Search?. In *Proceedings of iConference 2015*. iDEALS.

[15] Toine Bogers and Vivien Petras. 2017. An In-depth Analysis of Tags and Controlled Vocabulary for Book Search. In *Proceedings of iConference 2017*. iDEALS.

[16] P. Borlund and N. Pharo. 2019. A Need for Information on Information Needs. *Information Research, Proceedings of CoLIS'19* 24, 4 (2019).

[17] Andrei Broder. 2002. A Taxonomy of Web Search. *ACM SIGIR Forum* 36, 2 (2002), 3–10.

[18] Roger Brown and David McNeill. 1966. The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior* 5, 4 (1966), 325–337. https://doi.org/10.1016/S0022-5371(66)80040-3

[19] H. Bruce. 2005. Personal, anticipated information need. *Information Research* 10, 3 (2005).

[20] Katriina Byström and Sanna Kumpulainen. 2020. Vertical and Horizontal Relationships amongst Task-based Information Needs. *Information Processing & Management* 57, 2 (2020), 102065.

[21] Robert G. Capra. 2006. *An Investigation of Finding and Refinding Information on the Web*. Ph.D. Dissertation. Virginia Polytechnic Institute and State University.

[22] Suzanne Chapman, Shevon Desai, Kat Hagedorn, Ken Varnum, Sonali Mishra, and Julie Piacentine. 2013. Manually Classifying User Search Queries on an Academic Library Web Site. *Journal of Web Librarianship* 7, 4 (2013), 401–421.

[23] Charles Cole. 2011. A Theory of Information Need for Information Retrieval that Connects Information to Knowledge. *Journal of the American Society for Information Science and Technology* 62, 7 (2011), 1216–1231.

[24] Mats Dahlström and Mikael Gunnarsson. 2000. Document Architecture Draws a Circle: On Document Architecture and its Relation to Library and Information Science Education and Research. *Information Resarch* 5, 2 (2000).

[25] David Elsweiler, Mark Baillie, and Ian Ruthven. 2008. Exploring memory in email refinding. *ACM Transactions on Information Systems* 26, 4 (2008), 21.

[26] David Elsweiler, Mark Baillie, and Ian Ruthven. 2011. What Makes Re-finding Information Difficult? A Study of Email Re-finding. In *Advances in Information Retrieval*, Paul Clough, Colum Foley, Cathal Gurrin, Gareth J. F. Jones, Wessel Kraaij, Hyowon Lee, and Vanessa Mudoch (Eds.). Springer Berlin Heidelberg, 568–579.

[27] David Elsweiler and Ian Ruthven. 2007. Towards Task-based Personal Information Management Evaluations. In *SIGIR '07: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, USA, 23–30.

[28] David Elsweiler, Max L. Wilson, and Brian Kirkegaard Lunn. 2011. *Understanding Casual-Leisure Information Behaviour*. Emerald, Bingley, West Yorkshire, 211–241.

[29] J.L. Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin* 76, 5 (1971), 378–382.

[30] Maria Gäde, Mark Michael Hall, Hugo C. Huurdeman, Jaap Kamps, Marijn Koolen, Mette Skov, Elaine Toms, and David Walsh. 2015. First Workshop on Supporting Complex Search Tasks. In *Proceedings of the First International Workshop on*

Supporting Complex Search Tasks co-located with the 37th European Conference on Information Retrieval (ECIR 2015), Vienna, Austria, March 29, 2015.

[31] Matthias Hagen, Daniel Wägner, and Benno Stein. 2015. A Corpus of Realistic Known-Item Topics with Associated Web Pages in the ClueWeb09. In *ECIR '15: Proceedings of the 37th European Conference on Information Retrieval*. Springer, 513–525.

[32] Morgan Harvey and David Elsweiler. 2012. Exploring Query Patterns in Email Search. In *Advances in Information Retrieval*, Ricardo Baeza-Yates, Arjen P. de Vries, Hugo Zaragoza, B. Barla Cambazoglu, Vanessa Murdock, Ronny Lempel, and Fabrizio Silvestri (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 25–36.

[33] Peter Ingwersen and Kalervo Järvelin. 2005. *The Turn: Integration of Information Seeking and Retrieval in Context*. Springer.

[34] Bernard J. Jansen, Danielle L Booth, and Amanda Spink. 2008. Determining the Informational, Navigational, and Transactional Intent of Web Queries. *Information Processing & Management* 44, 3 (2008), 1251–1266.

[35] W. Jones and J. Teevan. 2007. How Do People Find Personal Information. In *Personal Information Management*, W. Jones and J. Teevan (Eds.). University of Washington Press, Washington, 22–34.

[36] Ida Kathrine Hammeleff Jørgensen and Toine Bogers. 2020. "Kinda like The Sims... But with ghosts?": A Qualitative Analysis of Video Game Re-finding Requests on Reddit. In *FDG '20: Proceedings of the 15th International Conference on the Foundations of Digital Games*. ACM.

[37] Ja-Young Kim and Nicholas J Belkin. 2002. Categories of Music Description and Search Terms and Phrases Used by Non-Music Experts.. In *Proceedings of ISMIR '02*, Vol. 2. 209–214.

[38] Marijn Koolen. 2014. User Reviews in the Search Index? That'll Never Work!. In *ECIR '14: Proceedings of the 36th European Conference on Information Retrieval*. Springer, 323–334.

[39] Marijn Koolen, Toine Bogers, Jaap Kamps, and Antal van den Bosch. 2015. Looking for Books in Social Media: An Analysis of Complex Search Requests. In *ECIR '15: Proceedings of the 37th European Conference on Information Retrieval (Lecture Notes in Computer Science, Vol. 9022)*. Springer, 184–196.

[40] Marijn Koolen, Jaap Kamps, Toine Bogers, Nicholas J. Belkin, Diane Kelly, and Emine Yilmaz. 2017. Current Research in Supporting Complex Search Tasks. In *Proceedings of the Second Workshop on Supporting Complex Search Tasks co-located with the ACM SIGIR Conference on Human Information Interaction & Retrieval (CHIIR 2017), Oslo, Norway, March 11, 2017*. 1–4.

[41] Marijn Koolen, Jaap Kamps, and Gabriella Kazai. 2012. Social Book Search: Comparing Topical Relevance Judgements and Book Suggestions for Evaluation. In *CIKM '12: Proceedings of the 21st ACM International Conference on Information and Knowledge Management*. ACM, 185–194.

[42] J. Richard Landis and Gary G. Koch. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics* (1977), 159–174.

[43] Jin Ha Lee and J Stephen Downie. 2004. Survey of Music Information Needs, Uses, and Seeking Behaviours: Preliminary Findings.. In *ISMIR*, Vol. 2004. Citeseer, 5th.

[44] Jin Ha Lee, Allen Renear, and Linda C Smith. 2006. Known-Item Search: Variations on a Concept. *ASIST '06: Proceedings of the American Society for Information Science and Technology* 43, 1 (2006), 1–17.

[45] Thomas W. Malone. 1983. How Do People Organize Their Desks?: Implications for the Design of Office Information Systems. *ACM Transactions on Information Systems* 1, 1 (1983), 99–112.

[46] Florian Meier and David Elsweiler. 2014. Tweets I've Seen: Analysing Factors Influencing Re-Finding Frustration on Twitter. In *Proceedings of the 5th Information Interaction in Context Symposium* (Regensburg, Germany) *(IIiX '14)*. ACM, New York, NY, USA, 287–290.

[47] Florian Meier and David Elsweiler. 2016. Going Back in Time: An Investigation of Social Media Re-finding. In *Proceedings of SIGIR '16* (Pisa, Italy). ACM, New York, NY, USA, 355–364.

[48] Meredith Ringel Morris. 2008. A Survey of Collaborative Web Search Practices. In *CHI '08: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1657–1660.

[49] Hartmut Obendorf, Harald Weinreich, Eelco Herder, and Matthias Mayer. 2007. Web Page Revisitation Revisited: Implications of a Long-Term Click-Stream Study of Browser Usage. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '07)*. ACM, New York, NY, USA, 597–606.

[50] Paul Ogilvie and Jamie Callan. 2003. Combining Document Representations for Known-Item Search. In *SIGIR '03: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, USA, 143–150.

[51] Greg Pass, Abdur Chowdhury, and Cayley Torgeson. 2006. A Picture of Search. In *InfoScale '06: The First International Conference on Scalable Information Systems*.

[52] Kara Reuter. 2007. Assessing Aesthetic Relevance: Children's Book Selection in a Digital Library. *JASIST* 58, 12 (2007), 1745–1763.

[53] Kerry Rodden and Kenneth R. Wood. 2003. How Do People Manage Their Digital Photographs?. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI '03)*. ACM, New York, NY, USA, 409–416.

[54] Ian Ruthven. 2019. The Language of Information Need: Differentiating Conscious and Formalized Information Needs. *Information Processing & Management* 56, 1 (2019), 77–90.

[55] Seyedeh Sargol Sadeghi, Roi Blanco, Peter Mika, Mark Sanderson, Falk Scholer, and David Vallet. 2017. Re-Finding Behaviour in Vertical Domains. *ACM Transactions on Information Systems* 35, 3 (2017), 21.

[56] Tefko Saracevic. 1996. Relevance Reconsidered. In *CoLIS 2: Proceedings of the 2nd Conference on Conceptions of Library and Information Science*. ACM, 201–218.

[57] Reijo Savolainen. 1995. Everyday Life Information Seeking: Approaching Information Seeking in the Context of "Way of Life". *Library & Information Science Research* 17, 3 (1995), 259 – 294.

[58] R. Savolainen. 2017. Information Need as Trigger and Driver of Information Seeking: A Conceptual Analysis. *Aslib Journal of Information Management* 69, 1 (2017), 2–21.

[59] Margrit Schreier. 2014. Ways of Doing Qualitative Content Analysis: Disentangling Terms and Terminologies. *Forum: Qualitative Social Research* 15, 1 (2014).

[60] Ian Soboroff. 2004. On Evaluating Web Search with Very Few Relevant Documents. In *SIGIR '04: Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, USA, 530–531.

[61] Tyler Tate and Tony Russell-Rose. 2012. The Information Needs of Mobile Searchers: A Framework. In *Proceeding of Searching4Fun Workshop*.

[62] Robert S Taylor. 1968. Question-Negotiation and Information Seeking in Libraries. *College and Research Libraries* 29 (1968), 178–194.

[63] Jaime Teevan, Eytan Adar, Rosie Jones, and Michael A. S. Potts. 2007. Information Re-Retrieval: Repeat Queries in Yahoo's Logs. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Amsterdam, The Netherlands) *(SIGIR '07)*. ACM, New York, NY, USA, 151–158.

[64] Johanne R. Trippas, Paul Thomas, Damiano Spina, and Hideo Joho. 2020. CAIR'20: Third International Workshop on Conversational Approaches to Information Retrieval, co-located at CHIIR 2020. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval* (Vancouver BC, Canada) *(CHIIR '20)*. ACM, New York, NY, USA, 492–494.

[65] Alexandra Vtyurina and Adam Fourney. 2018. Exploring the Role of Conversational Cues in Guided Task Support with Virtual Assistants. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. ACM, New York, NY, USA, 1–7.

[66] Liu Yang, Susan T. Dumais, Paul N. Bennett, and Ahmed Hassan Awadallah. 2017. Characterizing and Predicting Enterprise Email Reply Behavior. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Shinjuku, Tokyo, Japan) *(SIGIR '17)*. ACM, New York, NY, USA, 235–244.

[67] Hamed Zamani and W. Bruce Croft. 2020. Learning a Joint Search and Recommendation Model from User-Item Interactions. In *WSDM '20: Proceedings of the 13th ACM International Conference on Web Search and Data Mining*. ACM, New York, NY, USA.